New Results on Universal Dynamic Regret Minimization for Learning and Control

Yu-Xiang Wang

Based on joint work with --------------->

Dheeraj Baby
(on the job market)

COMPUTER SCIENCE
UC SANTA BARBARA
Computing. ReInvented.

Statistical Machine Learning Research at UCSB

- Differential Privacy, Private Learning
- Offline and Low-Adaptive Reinforcement Learning
- Adaptive Online Learning
- Theory of Deep Learning

Ming Yin
(Also visiting!)

UCSB ML Lab

Center for Responsible Machine Learning

Our research is partially supported by:

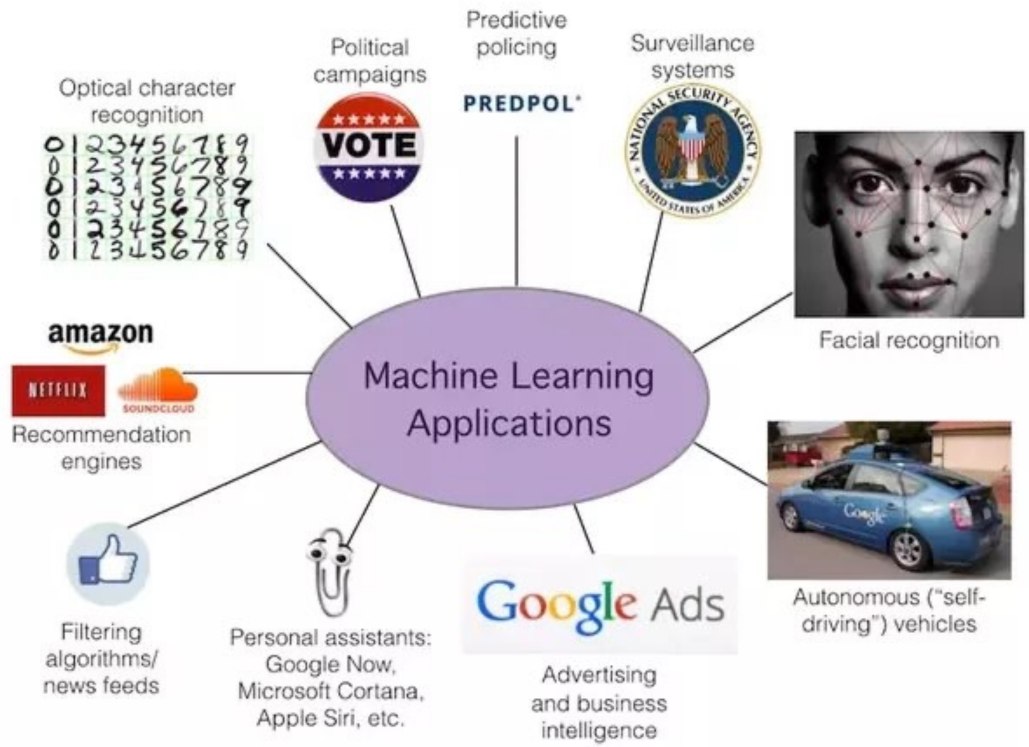
NSF, aws, Adobe, Google, evidation, NEC, appfolio

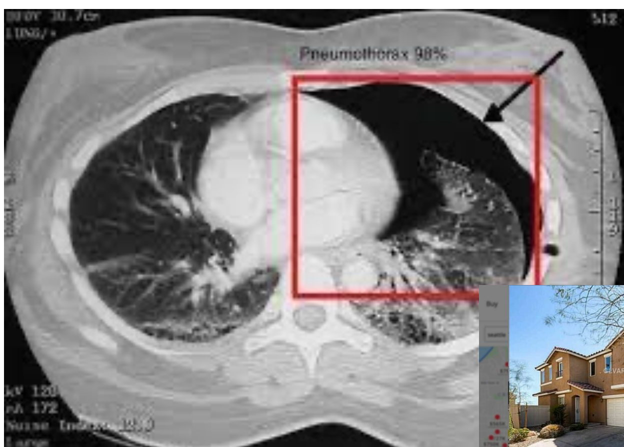
Positions available: PhD students, postdoc, new faculty, sabbatical visitor

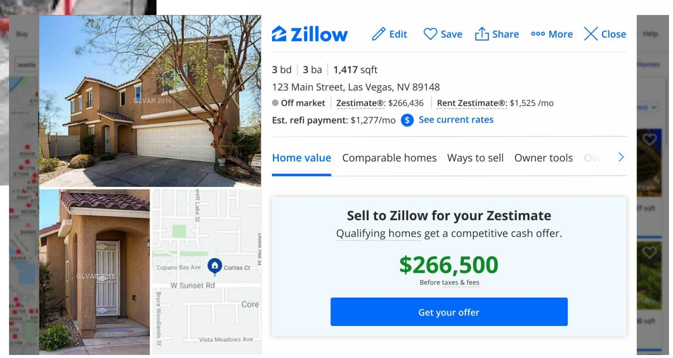
Outline

- Universal Dynamic Regret in online learning
  - Motivation and application
  - New results for *curved* loss functions
- Optimal Universal Dynamic Regret
  - Lower bound via non-parametric regression
  - Algorithm and proof sketch
  - From improper to proper learning
- Optimal Dynamic Regret in LQR Control
- Open problems / future work

~~AI~~ Machine Learning has revolutionized almost every aspect of our daily life







Most theory of ML relies on stochastic assumptions on data-generating processes

- **Parametric / Bayesian methods:** model the data generating process up to some parameters

- **Nonparametric statistics:** Consider *very broad families* of distributions where the data can be coming from.

- **Statistical Learning Theory:**
  - Assume data drawn iid (from any distribution)

What if the data are not drawn iid or even stochastic?

Online learning --- a powerful learning paradigm that makes no stochastic assumptions

The Online Learning setting

- For each $t \in [n] := \{1, \ldots, n\}$, learner predicts $\boldsymbol{x}_t \in \mathcal{D} \subset \mathbb{R}^d$.
- Adversary reveals a loss function $f_t : \mathbb{R}^d \to \mathbb{R}$

Example: $f_t(x) = (\mathrm{StockPrice}_t - \mathrm{Feature}_t^T x)^2$

(Static) Regret: Compete with any fixed $\mathbf{w} \in \mathcal{W} \subset \mathcal{D}$ chosen in hindsight:

$$R_n(\mathbf{w}) := \sum_{t=1}^{n} f_t(\mathbf{x}_t) - \sum_{t=1}^{n} f_t(\mathbf{w})$$

- Excellent treatment on this subject by Vovk, Lugosi, Ceca-Bianchi, Hazan, Shalev-Schwartz, Orabona et al…

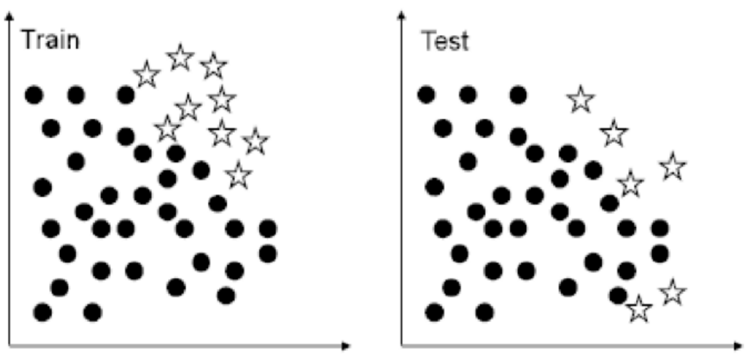
Well-known results on no-regret online learning

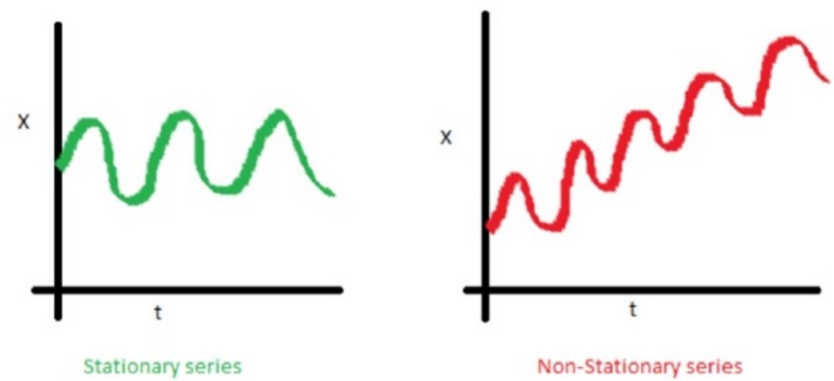
	Optimal regret
Convex losses	$\Theta(\sqrt{n})$
Strongly convex losses	$\Theta(\log n)$
Exp-concave losses	$\Theta(\log n)$

* various problem parameters omitted for simplicity.

- Excellent treatment on this subject by Vovk, Lugosi, Ceca-Bianchi, Hazan, Shalev-Schwartz, Orabona et al…

Nonstationarity “Change is the only constant in life”





- Viruses mutate. A drug that passes a clinical trial in 2020 may become ineffective in 2021.
- Trendy topics change over time. Language models trained on older data may struggle to remain relevant.
- Stock prices are affected by events. A trading strategy can work amazingly well in one period but fail miserably when market condition changes.

Static Regret Bound is not so useful in nonstationary environments

Tesla stock price 2011-2021

(taken from Reddit)

Best linear prediction in hindsight

Can we handle nonstationarity without modeling the world? Yes, by Universal Dynamic Regret Minimization

The Online Learning setting

- For each $t \in [n] := \{1, \ldots, n\}$, learner predicts $\boldsymbol{x}_t \in \mathcal{D} \subset \mathbb{R}^d$.
- Adversary reveals a loss function $f_t : \mathbb{R}^d \to \mathbb{R}$

Example: $f_t(x) = (\mathrm{StockPrice}_t - \mathrm{Feature}_t^T x)^2$

Goal: Learner aims to control its dynamic regret against any sequence of comparators $\boldsymbol{w}_1, \ldots \boldsymbol{w}_n$ where $\boldsymbol{w}_t \in \mathcal{W} \subseteq \mathcal{D}$ for all t.

$$R_n(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n) := \sum_{t=1}^{n} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{w}_t),$$

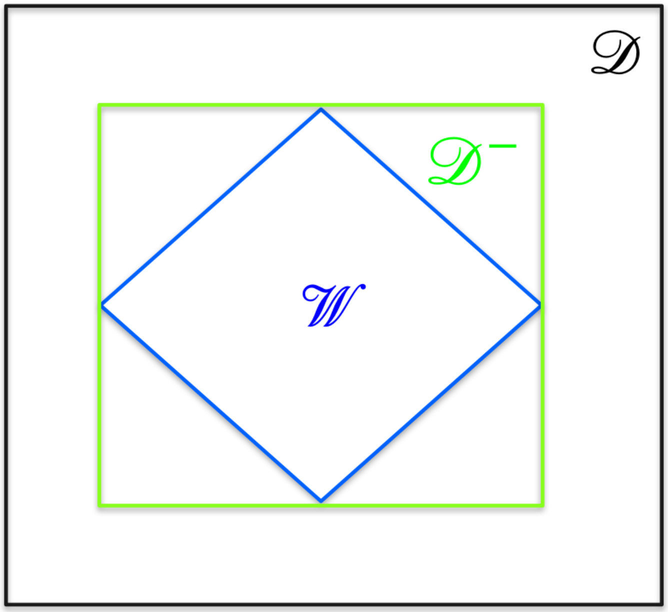
Proper learning vs improper learning

- Benchmark set $\mathcal{W}$, Decision set $\mathcal{D}$

Learning is said to be proper when $\mathcal{W} = \mathcal{D}$.

Learning is said to be improper when $\mathcal{W} \subset \mathcal{D}$.

- In the first part our our results we consider improper learning



Several burning questions

1. How does this address non-stationarity?

   More on this after seeing the results!

2. The worst-case dynamic regret seems linear?

3. Why qualifying it with “Universal”?

4. What are your new results?

5. Connections to “adaptive regret”?

   Later, in the proof!

Dynamic regret is parameterized by properties of each comparator sequence.

$$R_n(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n) := \sum_{t=1}^{n} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{w}_t),$$

- Worst-case dynamic regret is linear.
- Often parameterized by how much the comparator sequence changes over time, i.e., total variation.

$$P_n(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n) = \sum_{t=1}^{n} \|\boldsymbol{w}_t - \boldsymbol{w}_{t-1}\|_2$$

$$C_n(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n) = \sum_{t=1}^{n} \|\boldsymbol{w}_t - \boldsymbol{w}_{t-1}\|_1$$

Why "universal"? Because we want to *simultaneously* compete with *all* comparator sequences

- It implies an "Oracle Inequality"

$$\underbrace{\sum_{t=1}^{n} f_t(x_t)}_{\text{Our performance}} \le \min_{w_1,\ldots,w_n} \underbrace{\sum_{t=1}^{n} f_t(w_t)}_{\text{Comparator performance}} + \underbrace{\mathrm{RegretBound}(w_{1:n})}_{\text{Dynamic regret}}$$

- This is in contrast to the "restricted dynamic regret"

$$\sum_{t=1}^{n} f_t(x_t) - \sum_{t=1}^{n} f_t(w_t^*) \qquad \text{where } w_t^* = \mathrm{argmin}_w f_t(w)$$

*The restricted version were considered in (Besbes et al, 2013) (Jadbadie et al., 2016) under different feedback models.

Universal vs Restrictive Dynamic Regret in Online Linear Regression

- Example:

$$f_t(x) = (\mathrm{StockPrice}_t - \mathrm{Feature}_t^T x)^2$$

- Restrictive Dynamic regret competes with an unrealistic oracle that achieves 0 loss, but incur O(n) regret

- Universal dynamic regret competes with more stable policies with sublinear loss and regret
  - Fundamental values change slowly
  - Optimal bias-variance tradeoff

Existing results on dynamic regret minimization since Zinkevich (2003)

	Static regret	Dynamic Regret bounds
Convex losses	$\Theta(\sqrt{n})$	$\Theta(\sqrt{n(1+C_n)})$
Strongly convex losses	$\Theta(\log n)$	Open problem
Exp-concave losses	$\Theta(\log n)$	

The case with unknown C_n was resolved in (Zhang and Zhou, 2018).

Only minor improvement known (Yuan and Lamperski, 2019) $\tilde{O}(1+\sqrt{nC_n})$

* various problem parameters omitted for simplicity.

Fast rates under exp-concave losses are useful. Many useful applications satisfy exp-concavity!

- Example 1: Online Nonparametric Regression

$$f_t(x_t) = (y_t - x_t)^2 \quad \text{where } y_t = \theta_t + \mathcal{N}(0, \sigma^2)$$

- Example 2: Online Linear Regression

$$f_t(x) = (\mathrm{StockPrice}_t - \mathrm{Feature}_t^T x)^2$$

- Example 3: Online Logistic Regression

$$f_t(x) = \log(1 + e^{-y_t \mathrm{Feature}_t^T x}) \qquad y_t \in \{-1, 1\}$$

- Example 4: Universal Portfolio Selection

$$f_t(x) = \log(x^T r_t)$$

*weak assumption e.g., boundedness often needed.

Summary of our new results


Dheeraj Baby

	Static regret	**Dynamic Regret bounds**
Convex losses	$\Theta(\sqrt{n})$	$\Theta(\sqrt{n(1+C_n)})$
Strongly convex losses	$\Theta(\log n)$	$\tilde{\Theta}(n^{1/3}C_n^{2/3} \vee 1)$
Exp-concave losses	$\Theta(\log n)$	$\tilde{\Theta}(n^{1/3}C_n^{2/3} \vee 1)$

* various problem parameters omitted for simplicity.

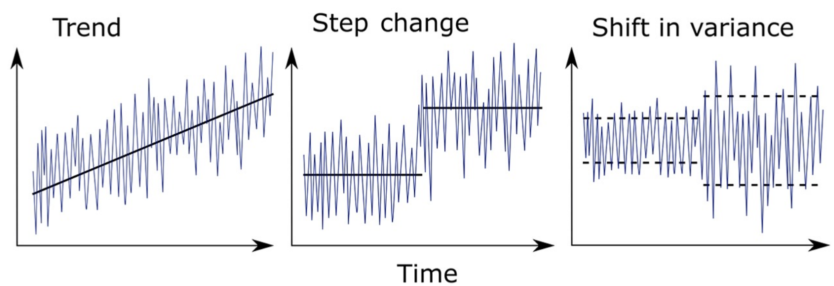
(Baby and W., COLT'21 Best Student Paper)
- Improper learning.
- need smoothness
- extra d dependence

(Baby and W., AISTATS'22)
+ Proper learning
+ No smoothness & Optimal dim dep for strongly convex cases.
- only for box constraints for exp-concave losses

Now how does this address non-stationarity?

- It is fully agnostic and it does not make assumptions about the type of non-stationarity

Covariate Shift	**Label Shift**
$q(\boldsymbol{x}, y) = q(\boldsymbol{x})p(y\|\boldsymbol{x})$	$q(\boldsymbol{x}, y) = q(y)p(\boldsymbol{x}\|y)$



- Optimally compete with your favorite sequence chosen **in hindsight**

Outline

- Universal Dynamic Regret in online learning
    - Motivation and application
    - New results for *curved* loss functions

- Optimal Universal Dynamic Regret
    - Lower bound via non-parametric regression
    - Algorithm and proof sketch
    - From improper to proper learning

- Optimal Dynamic Regret in LQR Control

- Open problems / future work

Dynamic regret is parameterized by the total variation of the comparator sequence

Dheeraj Baby

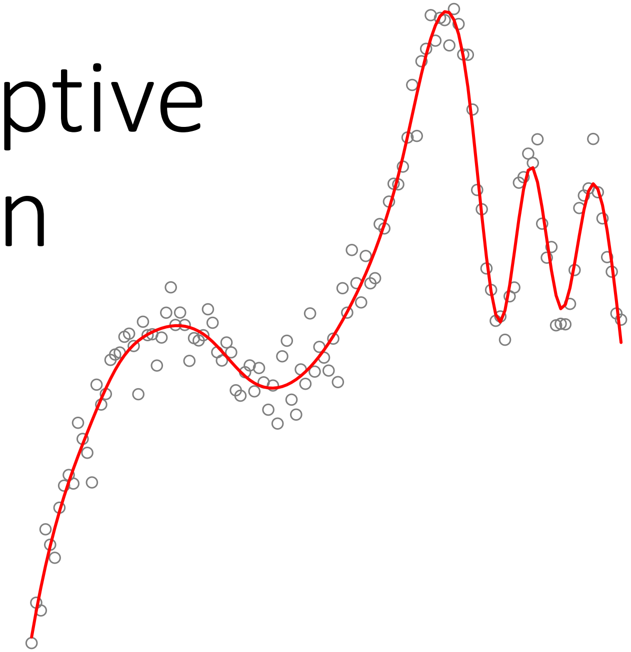
$$C_n(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n) = \sum_{t=1}^{n} \|\boldsymbol{w}_t - \boldsymbol{w}_{t-1}\|_1$$

Theorem 1 (simplified): (Baby and W., 2021)

For exp-concave and smooth losses, there is an efficient improper online algorithm, s.t.

$$\underbrace{\sum_{t=1}^{n} f_t(x_t)}_{\text{Our performance}} \leq \min_{w_1,\ldots,w_n} \underbrace{\sum_{t=1}^{n} f_t(w_t)}_{\text{Comparator performance}} + \underbrace{O(n^{1/3} C_n(w_1, \ldots, w_n)^{2/3})}_{\text{Dynamic regret}}$$

Connection to locally adaptive non-parametric regression



$$f_t(x_t) = (y_t - x_t)^2$$

where $y_t = \theta_t + \mathcal{N}(0, \sigma^2)$

Take $w_1, ..., w_n = \theta_1, ..., \theta_n$

$$\sum_{t=1}^{n} (y_t - x_t)^2 \leq \sum_t (y_t - \theta_t)^2 + O(n^{1/3} C_n(\theta_1, ..., \theta_n)^{2/3})$$

take expectation ⬇ **divide by n**

MSE $\mathbb{E}[\frac{1}{n}\sum_{t=1}^{n}(x_t - \theta_t)^2] = O(n^{-2/3} C_n(\theta_1, ..., \theta_n)^{2/3})$

Optimal rate for estimating functions in TV class!

It is more flexible than the standard nonparametric regression, because

- No statistical assumptions

- No hyperparameter / adaptively optimal

- More general loss functions

- Can be used for online forecasting

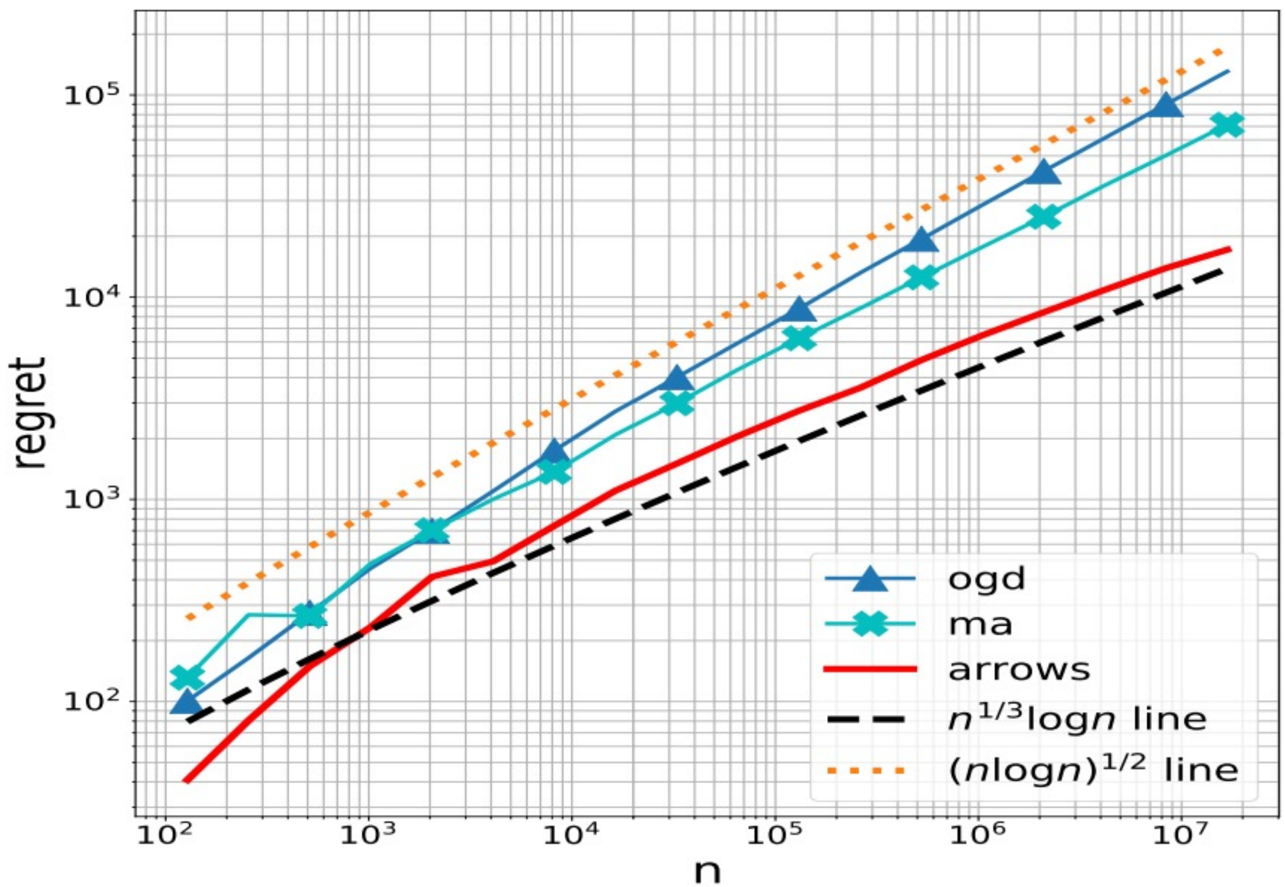
Another interesting lower bound from nonparametric regression

- OGD with *any* learning rate schedule are **Linear Estimators** for the non-parametric regression problem. (Baby and W., NeurIPS'19)

- The lower bound by Donoho, Liu, MacGibbon (1990) => Restarting-OGD and Ader require $\Omega\left(\sqrt{nC_n}\right)$ regret!

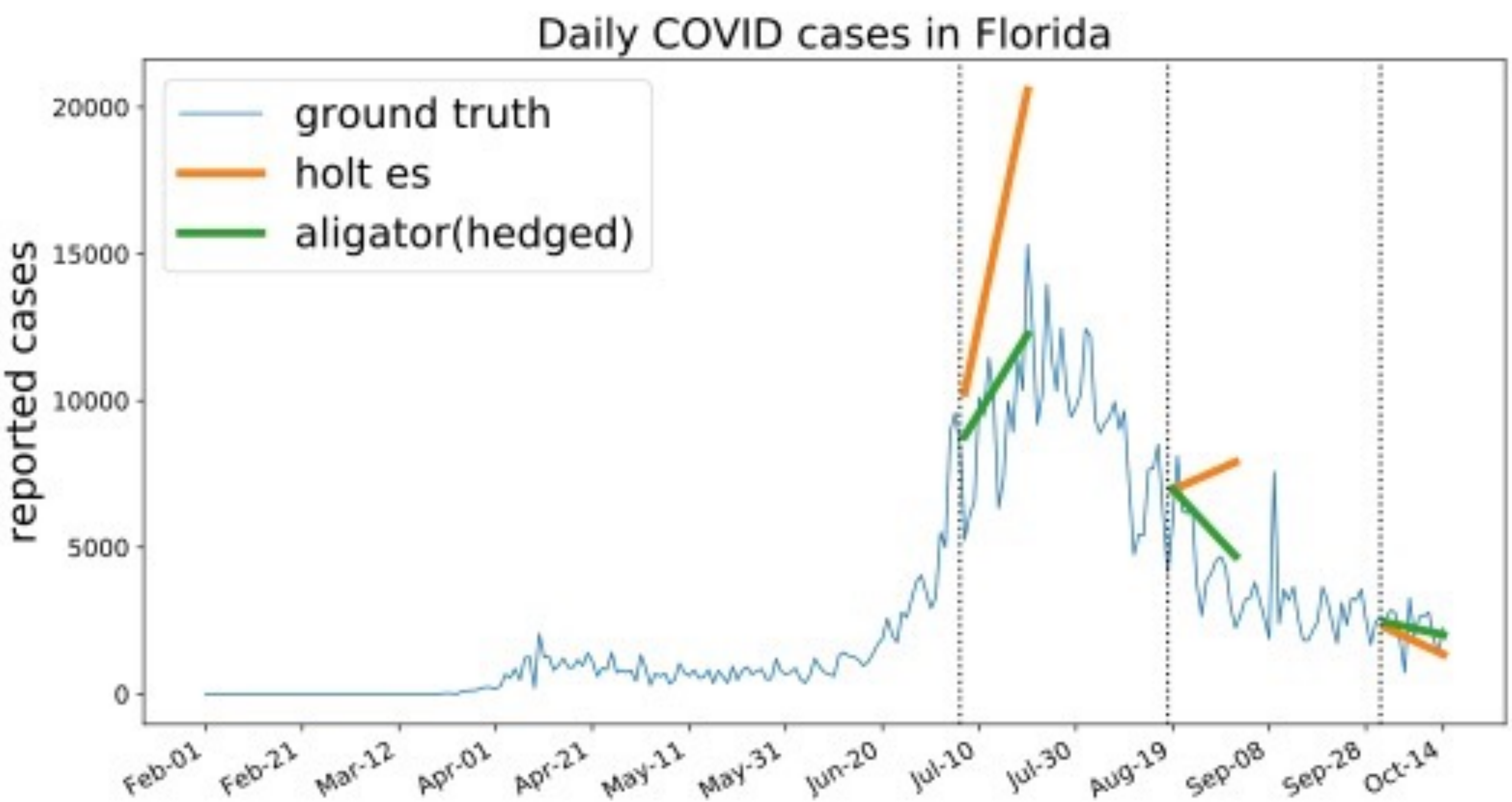
- Cannot achieve the optimal $n^{1/3}C_n^{2/3}$
  - Those methods that achieve this rate is known as **"locally adaptive"** methods, e.g., wavelets, adaptive kernels, adaptive splines, trend filtering etc.

Separation of non-adaptive and adaptive methods on this problem, numerically…



(Baby and W., 19) Online Forecasting of TV-bounded sequences

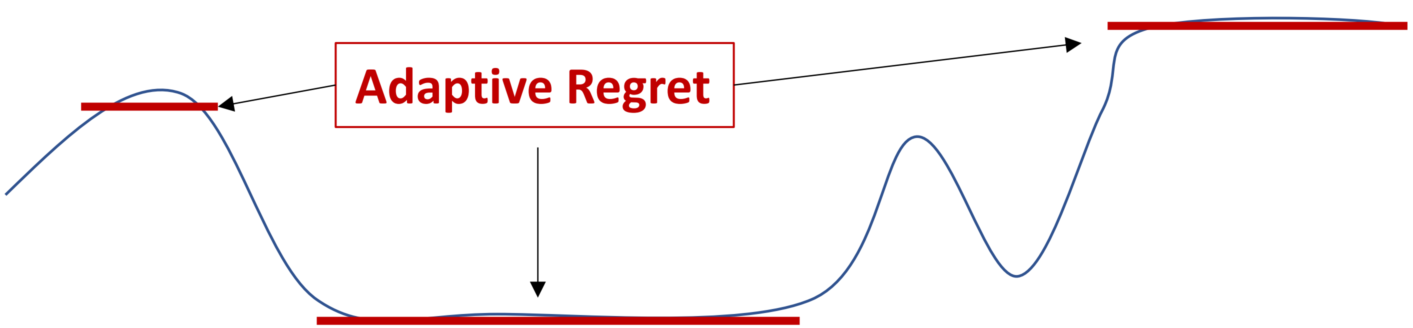
Application to "Online Trend Removal" in COVID hospitalization forecasting



(Baby, Zhao and W., AISTATS'21)

Proof highlights: Adaptive Regret and Strongly Adaptive Online Learner

- Adaptive Regret Minimization (Hazan and Seshadhri, 2009) (Daniely, Gonen, Shalev-Shwartz, 2015)
  - Follow the Leading History (FLH)
  - (Essentially) running multiplicative weights over an ensemble of Online Learners that starts at every time step.
- Our algorithm: FLH with Online-Newton-Step
- For exp-concave losses, FLH-ONS achieves an $\tilde{O}(1)$ **static regret** of on **all intervals** at the same time!



Proof highlights: (TV-Constrained) Offline Optimal Comparator

- For each $C_n \geq 0$, the **offline optimal** is solution to

$$\boxed{\begin{aligned} &\min_{u_{1:n}} \sum_{t=1}^{n} f_t(u_t) \\ &\text{s.t.} \quad \sum_{t=2}^{n} |u_t - u_{t-1}| \leq C_n \\ &\qquad\quad |u_t| \leq B \text{ for } t = 1, 2, ..., n \end{aligned}}$$

- It suffices to bound the dynamic regret against offline optimal for each C_n

$$R_n(w_{1:n}) \leq R_n(u^*_{1:n})$$

  - For all $w_{1:n}$ satisfying the total variation bound C_n

Proof highlights: KKT conditions of the offline optimal

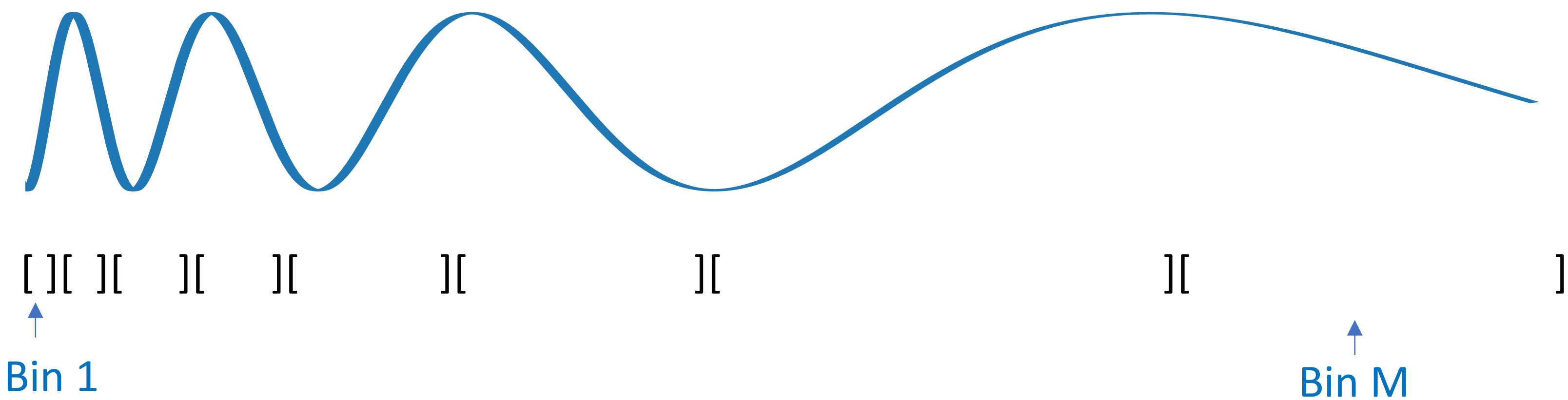
- The solution is somewhat special in that it satisfies a set of KKT conditions.

- ***stationarity:*** $\nabla f_t(u_t) = \lambda\left(s_t - s_{t-1}\right) + \gamma_t^- - \gamma_t^+;$
- ***complementary slackness:*** *(a)* $\lambda\left(\sum_{t=2}^{n}|u_t - u_{t-1}| - C_n\right) = 0;$
  (b) $\gamma_t^-\left(u_t + B\right) = 0$ *and* $\gamma_t^+\left(u_t - B\right) = 0.$

- Gives rise to interesting combinatorial properties

Proof highlights: Adaptive Partition

- Let the following be the offline optimal comparator



We construct a partitioning of $[n]$ into M bins as follows $\{[1_s, 1_t], \ldots, [i_s, i_t], \ldots, [M_s, M_t]\}$ satisfying:

- $C_i := \sum_{j=i_s}^{i_t-1} |u_{j+1} - u_j| \leq B/\sqrt{n_i}$ where $n_i := i_t - i_s + 1$, $i \in [M]$.
- Number of bins obeys $M = O(n^{1/3} C_n^{2/3} B^{-2/3} \vee 1)$.

Suffices to prove the **dynamic regret** in each bin is $\tilde{O}(1)$.

Proof highlights: Regret Decomposition

One-step Gradient Descent

$$R_n(C_n) \le \sum_{i=1}^{M} \underbrace{\sum_{t=i_s}^{i_t} f_t(x_t) - f_t\left(\boxed{\bar{u}_i - \eta\nabla \sum_{t'=i_s}^{i_t} f_{t'}(\bar{u}_i)}\right)}_{T_{1,i}}$$

By Strong Adaptivity $T_{1,i} = O(B^2 \log n)$.

$$+ \sum_{i=1}^{M} \sum_{t=i_s}^{i_t} \underbrace{f_t\left(\bar{u}_i - \eta\nabla \sum_{t'=i_s}^{i_t} f_{t'}(\bar{u}_i)\right) - f_t(\bar{u}_i)}_{T_{2,i}}$$

By **Descent Lemma** $T_{2,i} \le -\frac{\eta}{2}\|\nabla\|^2$

$$+ \underbrace{\sum_{i=1}^{M} \sum_{t=i_s}^{i_t} f_t(\bar{u}_i) - f_t(u_t)}_{T_{3,i}}$$

By KKT conditions

$$\begin{aligned} T_{3,i} &\le n_i C_i^2 + 3\lambda C_i \\ &\le B^2 + 3\lambda C_i, \end{aligned}$$

* $T_{2,i}$ is not always strictly negative. $T_{3,i}$ is often very large. Turns out that there is a **magical refinement of the partition** such that $T_{2,i}$ is sufficiently negative when we need it be.

** The first time KKT conditions across time-steps are exploited in online learning.

A flavor of the splitting rules (for the square loss cases, without boundedness constraints)

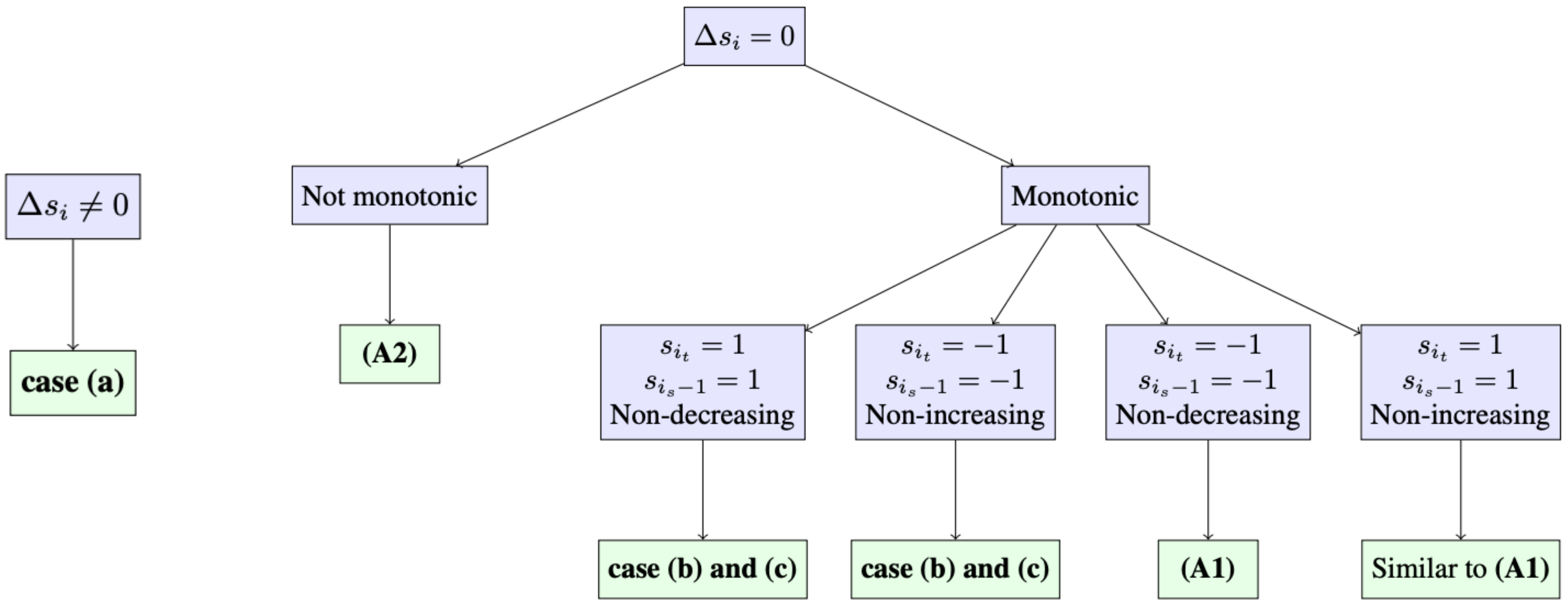


Figure 3: *Various configurations of the optimal sequence within a bin* $[i_s, i_t]$ *with* $\Delta s_i = 0$. *The leaf nodes indicate the labels of the paragraphs in the Proof of Theorem 1 to handle each scenario.*

- Case (a) (b) (c) can be directly bounded.
- Case (A1) (A2) can be converted into case(a) (b) (c) while doubling # of bins.

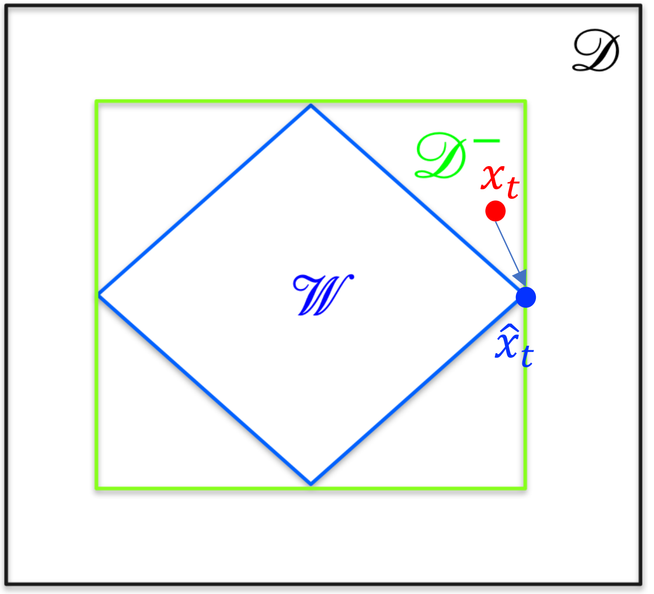
Challenges of proper learning

- The KKT conditions becomes more complex
- Projected one-step gradient descent does not provide sufficiently negative T_2
- Splitting of the bins becomes a lot more involved
- Turn out we can only solve for the cases **with box-constraints** --- one coordinate at a time.

Baby and W. (2022). Optimal dynamic regret in proper online learning with strongly convex losses and beyond.

Box constrained proper learner + a surrogate loss technique from (Cutkosky and Orabona, 2018) suffices to solve general proper learning!

- In each iteration:
  1. Get prediction x_t from box-constrained learner
  2. Play $\hat{\boldsymbol{x}}_t = \Pi_{\mathcal{W}}(\boldsymbol{x}_t) := \mathrm{argmin}_{\boldsymbol{y} \in \mathcal{W}} \|\boldsymbol{x}_t - \boldsymbol{y}\|_1$.
  3. Get loss f_t
  4. Construct surrogate loss $\ell_t(\boldsymbol{x}) = f_t(\boldsymbol{x}) + G \cdot S(\boldsymbol{x})$, where $S(\boldsymbol{x}) := \|\boldsymbol{x} - \Pi_{\mathcal{W}}(\boldsymbol{x})\|_1$.
  5. Pass the surrogate loss to the box-constrained learner.

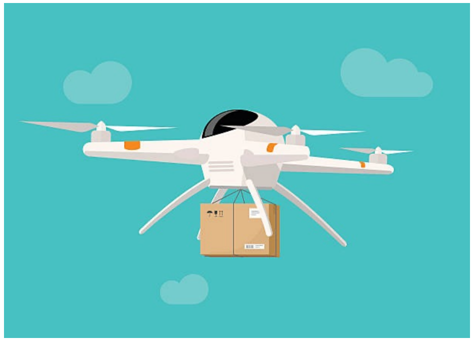


Adding $\|x_t - \hat{x}_t\|_1$ to the surrogate loss.

Outline

- Universal Dynamic Regret in online learning
  - Motivation and application
  - New results for *curved* loss functions
- Optimal Universal Dynamic Regret
  - Lower bound via non-parametric regression
  - Algorithm and proof sketch
  - From improper to proper learning
- Optimal Dynamic Regret in LQR Control
- Open problems / future work

Application to nonstochastic control



- Nonstochastic Control problem

$$x_{t+1} = Ax_t + Bu_t + w_t,$$

  - Linear dynamic systems, **known** dynamics.
  - Adversarial, but bounded disturbances
  - Existence of strongly stable controller

- Goal minimize the “dynamic policy regret”:

$$R\left(M_{1:n}\right) = \sum_{t=1}^{n} \ell(x_t^{\text{alg}}, u_t^{\text{alg}}) - \ell(x_t^{M_{1:n}}, u_t^{M_{1:n}}),$$

State-of-the-arts in online nonstochastic control problem

	General convex losses	LQR losses
Static Regret	$O(\sqrt{n})$ Reduction to OCO with Memory (Agarwal et al. 19)	$O(\log n)$ Reduction to online linear regression with delay (Foster and Simchowitz, 2020)
Universal Dynamic Regret	Dynamic regret version of OCO with memory $O(\sqrt{n(1+C_n)})$ (Zhao, W., and Zhou, 21)	Dynamic regret version of *Proper* online linear regression $\tilde{O}(n^{1/3}C_n^{2/3} \vee 1)$ (Baby and W., 2022b)

*C_n is the total variation of the parameters of an arbitrary sequence of Disturbance-Action Policies (DAPs).

Key technical challenge: Proper Learning in Online "minibatched" Linear Regression

- Loss function of interest

$$f_t(x) = \|A_t x - b_t\|^2$$

Not strongly convex!

- Key idea: a new min-max barrier in CO-style surrogate. Use box-constrained exp-concave result.

ProDR.control: Inputs - Decision set $\mathcal{D}$, $G > 0$

1. At round t, receive w_t from $\mathcal{A}$.
2. Receive co-variate matrix $A_t := [a_{t,1}, \ldots, a_{t,p}]^T$.
3. Play $\hat{w}_t \in \operatorname{argmin}_{x \in \mathcal{D}} \max_{i=1,\ldots,p} |a_{t,i}^T (x - w_t)|$.
4. Let $\ell_t(w) = f_t(w) + G \cdot S_t(w)$, where $f_t(w) = \|A_t w - b_t\|_2^2$ and $S_t(w) = \min_{x \in \mathcal{D}} \max_{i=1,\ldots,p} |a_{t,i}^T (x - w)|$.
5. Send $\ell_t(w)$ to $\mathcal{A}$.

Outline

- Universal Dynamic Regret in online learning
  - Motivation and application
  - New results for *curved* loss functions
- Optimal Universal Dynamic Regret
  - Lower bound via non-parametric regression
  - Algorithm and proof sketch
  - From improper to proper learning
- Optimal Dynamic Regret in LQR Control
- Open problems / future work

Summary of our techniques and new results on Universal Dynamic Regret

- Non-uniform bin splitting
- Regret decomposition
- KKT-conditions of offline optimal

Improper learning under Exp-Concave losses (Baby and W., 21)

+ Box-constraints proper learner
+ Refined analysis
+ CO's surrogate loss idea

Proper learning under Strongly convex losses (Baby and W., 22)

+ Min-max barrier

Proper learning under minibatch linear regression losses (Baby and W., 22a) => Apply to LQR control

Remaining Open problems

Open Problem 1: Proper learning for general exp-concave losses

- We know how to solve online generalized linear models
- We know how to solve online LQR-control.
- But universal portfolio remains out of reach…

Open Problem 2: Dimension-free bounds (RKHS)

- Dimension dependence required due to the L1 definition of TV.
- If we use the L2 version of Path Length, it could give dimension-free bounds.

A broader perspective on dynamic regret

$$\sum_{t=1}^{n} f_t(x_t) \leq \min_{w_1,\ldots,w_n} \sum_{t=1}^{n} f_t(w_t) + \mathrm{RegretBound}(w_{1:n})$$

- Why restricting to total variation / path length?
  - Higher-order smoothness?
  - Periodic sequences
  - Other recurring / switching patterns

NeurIPS’20 O(n^{1/(2k+3)}) higher-order case “Online Trend Filtering”

In the pipeline: O(n^1/5) universal dynamic regret for TV1 for exp-concave losses in full adversarial setting

Adaptive online learning in statistical methodology

- Apply dynamic regret / adaptive regret machinery to more statistical problems.
  - Provable guarantees without stochastic assumptions
  - Works with general loss functions
  - Often free of hyperparameters / highly adaptive

Thank you for your attention!

References:

1. Baby and W. (2021) Optimal Dynamic Regret in Exp-Concave Online Learning
2. Baby and W. (2022a) Optimal Dynamic Regret in Proper Online Learning with Strongly Convex Losses and Beyond
3. Baby and W. (2022b) Optimal Dynamic Regret for LQR Control
4. Baby and W. (2022c) Second-Order Path Variational

Dheeraj Baby
(on the job market)